On semi-automated matching and integration of database schemas

Ünal Karakaş, Ö.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
## Contents

1 INTRODUCTION  
1.1 Motivation and Requirements Analysis  1  
1.2 Addressed Research Questions  7  
1.3 Objectives and Contributions of the Thesis  9  
1.4 Scope of the Research  10  
1.5 Research Method  11  
1.6 Outline of the Dissertation  12  

2 INTERLINKING AND INTEGRATING SCHEMAS - BACKGROUND  15  
2.1 Related Concepts  15  
2.2 Multidatabase Classification Based on Schema Coupling  20  
2.3 Schema Matching and Schema Integration  21  
2.4 Conclusion  27  

3 HETEROGENEITY  29  
3.1 Related Concepts  29  
3.2 Taxonomy of Heterogeneity Resulted Conflicts  30  
3.3 Challenges for Schema Matching  35  
3.4 Conclusion  39  

4 SASMINT APPROACH  41  
4.1 Related Research Approaches  41  
4.2 Proposed Approach: SASMINT  53  
4.3 Conclusion  92  

5 SASMINT DEVELOPMENT ARCHITECTURE  95  
5.1 Processing Steps of SASMINT  95  
5.2 Technologies Applied  95  
5.3 Main Components of the System  97  
5.4 How does the System Work?  97  
5.5 Conclusions  105
## Contents

6 EMPIRICAL VALIDATION OF SASMINT  107
  6.1 Schema Matching Evaluations in Related Research  107
  6.2 Quality Measures Used for Evaluating SASMINT  108
  6.3 Test Schemas  112
  6.4 Setup for the Experimental Evaluation  115
  6.5 Evaluation of Schema Matching–For "select all above threshold" strategy  116
  6.6 Evaluation of Schema Matching with Sampler  119
  6.7 Evaluation of Schema Integration Performance  125
  6.8 Conclusions  129

7 THESIS CONCLUSIONS AND FUTURE WORK  133
  7.1 Summary of General Approach  133
  7.2 Reflections on the Research Questions  134
  7.3 Future Work  136

A LIST OF AUTHOR’S PUBLICATIONS  139

B XSD FOR SDML  141

C CLASS DIAGRAM FOR SDML  145

D TEST SCHEMAS  149

E EVALUATION OF SCHEMA MATCHING – FOR “SELECT MAX ABOVE THRESHOLD” STRATEGY  159

F EVALUATION OF SCHEMA INTEGRATION-DETAILS OF STEPS  163

BIBLIOGRAPHY  167

SUMMARY  175

SAMENVATTING  177

ACKNOWLEDGMENTS  181